Facilitating Effective Collaboration under the JWCRP using the JASMIN Platform

Ag Stephens (ag.stephens@stfc.ac.uk). Science & Technology Facilities Council, Didcot, Oxon, OX11 0QX.

British Atmospheric Data Centre
NATIONAL CENTRE FOR ATMOSPHERIC SCIENCE
NATURAL ENVIRONMENT RESEARCH COUNCIL

Centre for Environmental Data Archival
SCIENCE AND TECHNOLOGY FACILITIES COUNCIL
NATURAL ENVIRONMENT RESEARCH COUNCIL

The JASMIN computing platform at NCAS BADC provides access to high-performance disk, a virtualisation platform, compute cluster and high-speed networking. One key aspect from the NCAS perspective is that resources are equally accessible to researchers in UK universities and the UK Met Office. This makes JASMIN an excellent candidate for hosting services to support collaborations under the Joint Weather and Climate Research Programme (JWCRP).

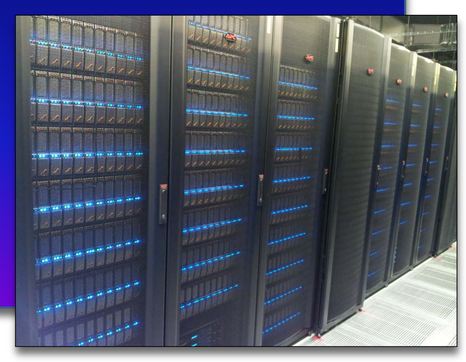
National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

Science & Technology Facilities Council

Joint Weather & Climate Research Programme
A partnership in weather and climate research

The Joint Weather & Climate Research Programme (**JWCRP**) is a shared programme between NERC and the Met Office. Its goals are to ensure that the UK maintains and strengthens its leading international position in weather and climate science, and hence in weather and climate forecasting and provision of advice for policy. The JWCRP involves a commitment to sustain and grow the UK's national capability and research in observing, understanding, modelling and predicting weather and climate, along with the infrastructure required to deliver that national capability and research. It also focuses on aligning major research initiatives and programmes to ensure the most effective impact of that research and pull-through into the delivery of services to government and business.

JASMIN is a data and computing platform designed to meet the needs of scientists working with very large datasets that require complex and computationally expensive analysis. Managed by STFC's Centre for Environmental Data Archival (CEDA) the platform now holds the complete archives of the NCAS BADC and NEODC. JASMIN also provides access to high-performance disk, a virtualisation environment, compute cluster and high-speed networking. Satellite nodes are deployed in the Universities of Leeds, Bristol and Reading. Plans to expand this capability are being developed.



WHAT CAN YOU DO ON JASMIN?

1. SHARE "LARGE" DATA WITH COLLABORATORS

Many projects are making JASMIN a focal point for collaboration by setting up a "Group Workspace". This is a shared disk made accessible to a chosen group of scientists working on the same project.

Group Workspaces currently **range between 5Tb and 380Tb** in size. Typically they are used for:

- Sharing large data – pushing from remote sites
- Accessing data – retrieving to remote sites
- Analysing data – space for storing big outputs
- Intercomparison – comparing data with established archived datasets, such as CMIP5 and ERA-Interim
- Overflow from HPC – staging outputs from costlier storage environments

2. ANALYSE DATA

Once logged into JASMIN, users can make the most of the **co-location with the NCAS BADC and CEMS Academic (NEODC) archives** as well as large Group Workspaces supporting specific projects. All data is stored on the high-performance storage platform.

Additionally, **generic scientific analysis virtual machines** are available for users to login and run code next to the data. The JASMIN Analysis Platform (Pascoe, 2013) is configured with a suite of **common software tools for data manipulation** (such as NetCDF libraries, NCO, CDO and the Python-Numpy stack).

3. EFFICIENT DATA TRANSFERS

JASMIN is built upon **4.6Pb (petabytes)** of PANASAS disk connected into the low latency network with 115 x 10Gbit/s connections. Workflows and data transfers can make use of the parallel networking to increase performance. Dedicated transfer servers are available for connection via SCP, RSYNC and GridFTP protocols. A 1Gbit/s "Lightpath" link is operational to the UK Met Office/MONSooN and a 2Gbit/s link to the HECTOR HPC site is currently being installed.

4. RUN YOUR OWN VMs (Root included!)

Specific projects may require dedicated computing resources and/or specific software environments. **JASMIN provides root access to project-specific virtual machines** so that external administrators can build and maintain specialised environments with minimum overhead and support. Projects, such as those discussed below (UPSCALE, NAME and PRECIS), typically marry this provision with a Group Workspace to store their inputs and outputs.

5. PARALLEL PROCESSING

The **"LOTUS" computing cluster** is available to users with processing needs that are computationally expensive. LOTUS is managed through batch queues which allow splitting of large jobs to run on many nodes or submission of model code that utilises parallel I/O functionality directly (such as OpenMPI). Centrally installed software is visible on all nodes and both archives and Group Workspaces can be accessed where permitted.

JWCRP-JASMIN HEADLINES

News from the first year...

UPSCALE - massive HPC outputs, transfers and analyses!

The NCAS/Met Office UPSCALE project relied upon JASMIN to act as a transfer node receiving over 200Tb of high-resolution climate simulations from the German HERMIT supercomputer. JASMIN support for the project included:

- Virtual machines for NCAS and Met Office
- 380Tb of Group Workspace storage and temporary backup (~200Tb)
- High-bandwidth allowed retrieval of 250Tb from Germany in 1 year
- Over 100Tb pulled back to Met Office archive

An UPSCALE spoke-scientist said, "We would never have been able to store, nor analyse, that volume of data, without the existence of the [JASMIN] service."
With analysis of outputs and the number of collaborators increasing, we expect the UPSCALE data to be relevant for years to come!

Atmospheric Dispersion Modelling

The Met Office atmospheric dispersion model, known as NAME, has been ported to JASMIN. Dedicated virtual machines will support a login service where the model can be run in a range of configurations driven by over 10Tb of meteorological data. This community service will also be coupled with a web-interface aimed at non-experts to calculate forward and back trajectories. See the NAME-on-JASMIN poster (Stephens *et al.*, 2013) for more information.

MONSooN HPC Overflow

Users of the NERC/Met Office MONSooN computing facility can transfer their data onto JASMIN where it can be analysed further, compared with other datasets, and archived in the long-term archive where appropriate. The recently installed 1Gbit dedicated "Lightpath" is available to MONSooN scientists wishing to migrate data quickly between the two sites.
Group Workspaces (shared disks for collaboration) can be set up for any MONSooN project to support NERC scientists managing and sharing their datasets.

Converting Climate runs to CF-NetCDF

Met Office software to convert the native PP-format into archive-ready CF-netCDF was deployed as the general purpose the "UM Conversion Service". Scientists from Reading and Edinburgh have used the tool to batch convert data. The latter submitted their outputs to the Paleoclimate Modelling Intercomparison Project 3 (PMIP3).

Regional Climate Modelling

The Met Office Hadley Centre's Regional Climate Modelling system (PRECIS) has been ported onto a virtual machine on JASMIN. The PRECIS team has been granted root access on the server, enabling them to install and configure the software and dependencies with minimum overhead. Over 30Tb of supporting data will be copied to a JASMIN Group Workspace to support running different configurations of PRECIS. A web-interface to certain configurations is also planned.

Refs & Links

B. N. Lawrence, V. Bennett, J. Churchill, M. Juckes, P. Kershaw, P. Oliver, M. Pritchard & A. Stephens (2012). *The JASMIN super-data-cluster*. http://arxiv.org/abs/1204.3553
Pascoe, S. (2013). *Growing a community analysis platform with JASMIN and the Community Intercomparison Suite*. NCAS Staff Meeting 2013. Poster.
Stephens, A. (2013). *Running the Met Office NAME dispersion model on the JASMIN computing platform: A new community tool and trajectory service*. NCAS Staff Meeting 2013. Poster.

UM Conversion Tool page on the CEDA Services wiki:
http://proj.badc.rl.ac.uk/cedaservices/wiki/JASMIN/SupportForScience/UMConversionTool
JASMIN website: http://www.jasmin.ac.uk
JWCRP website: http://www.jwcrp.org.uk
CEDA website: http://www.ceda.ac.uk